

On Modeling and Estimation for the Relative Risk and Risk Difference

THOMAS S. RICHARDSON

*Department of Statistics, University of Washington,
Box 354322, Seattle, Washington 98195, U.S.A.*

thomasr@uw.edu

JAMES M. ROBINS

*Department of Epidemiology, Harvard School of Public Health,
677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.*

robins@hsph.harvard.edu

LINBO WANG

*Department of Biostatistics, University of Washington,
Box 357232, Seattle, Washington 98195, U.S.A.*

lbwang@uw.edu

April 14, 2016

Abstract

A common problem in formulating models for the relative risk and risk difference is the variation dependence between these parameters and the baseline risk, which is a nuisance model. We address this problem by proposing the conditional log odds-product as a preferred nuisance model. This novel nuisance model facilitates maximum-likelihood estimation, but also permits doubly-robust estimation for the parameters of interest. Our approach is illustrated via simulations and a data analysis.

Keywords: Bivariate mapping; Estimating equation; Semi-parametric model; Variation independence

1 Introduction

The odds ratio (OR) is by far the most common way to model the association between binary random variables. The popularity of this measure is driven in part by the ease with which logistic regression can be used to describe the way that the odds ratio varies with baseline variables. However, there are fundamental difficulties that arise when attempting to interpret odds ratios. For example, even when treatment is randomized and hence independent of other baseline variables it is well known that the odds ratio is not collapsible, meaning that the marginal odds ratio will not lie in the convex hull of stratum-specific odds ratios (Rothman et al., 2008). This non-collapsibility makes it hard to interpret ORs and to compare logistic regression coefficients from different studies. The relative risk (RR) and the risk difference (RD) are alternative measures that are collapsible and are widely regarded as simpler to interpret. These are defined as follows:

$$\text{RR}(v) = \frac{P(Y=1 \mid A=1, V=v)}{P(Y=1 \mid A=0, V=v)},$$

$$\text{RD}(v) = P(Y=1 \mid A=1, V=v) - P(Y=1 \mid A=0, V=v),$$

where A is a binary exposure, Y is a binary response and V is a vector of covariates. See Lumley et al. (2006) for an extensive reference arguing for the use of RR/RD (or monotone transformations thereof) in place of ORs.

Any model for the conditional distribution $P(Y \mid A, V)$, such as a logistic regression, may be used to obtain an estimate for the RR or RD for a given covariate value v^* ; see Rothman et al. (2008, pp.439–440), McNutt et al. (2003). Although this approach does provide estimates for the RR or RD in any given stratum of V , nonetheless the parameters of a logistic model do not directly encode the dependence of the RR or RD on V . As a consequence the logistic model does not allow one to impose or test parsimonious models for the functional form of this dependence, nor does it offer direct insight into which baseline variables are important.

The current standard approach for modeling the dependence of RR or RD on baseline covariates assumes a Generalized Linear Model (GLM), under which there exist vectors ν , μ such that:

$$g\{E[Y \mid A, V]\} = A\nu^T W + \mu^T Z, \tag{1.1}$$

where $g(\cdot)$ is the link function, and $W = w(V)$, $Z = z(V)$ are known vector functions of V . With Y Bernoulli and g is the log link $g(u) = \log(u)$, equation (1.1) represents the mean function induced by a log-binomial (or Poisson) regression model; when g is the identity link $g(u) = u$, equation (1.1) represents the mean function induced by a linear probability (or linear regression) model.

With these choices for $g(\cdot)$, equation (1.1) implies, respectively, that $\log\{\text{RR}(V)\} = \nu^T W$ or $\text{RD}(V) = \nu^T W$. To see this, note that if $g = \log$ then (1.1) can be rewritten in the following *equivalent* form:

$$\log(\text{RR}(V)) = \nu^T W, \quad (1.2)$$

$$\log(p_0(V)) = \mu^T Z, \quad (1.3)$$

where $p_a(V) \equiv E[Y \mid A=a, V]$, $a=0, 1$. Similarly with the identity link and RD. Note that (1.2) models the parameter of interest, while (1.3) is a nuisance model used in estimating the parameter of interest.

However, these models present a new problem: the (joint) range of the left side of (1.2) and (1.3) is a constrained subspace in \mathbb{R}^2 . That is, $\text{RR}(v)$ (or $\text{RD}(v)$) and $p_0(v)$ are *variation dependent* in the sense that the range of $p_0(v)$ depends on the value of $\text{RR}(v)$ (or $\text{RD}(v)$). For example, if for some v^\dagger , $\text{RR}(v^\dagger) = 2$ so that $p_1(v^\dagger) = 2 \times p_0(v^\dagger)$, then clearly $p_0(v^\dagger) \leq 0.5$! Consequently, the range of $(\log\{\text{RR}(v)\}, \log\{p_0(v)\})$ or similarly, $(\text{RD}(v), p_0(v))$ is strictly smaller than the Cartesian product of their marginal ranges. Although one could easily find smooth monotone transformations of $\log\{\text{RR}(v)\}$ (or $\text{RD}(v)$) and $\log\{p_0(v)\}$ (or $p_0(v)$) that map each of their domains individually *onto* \mathbb{R} , nonetheless due to the variation dependence, the same issue remains: the joint range of the images of these transformations is still strictly smaller than the Cartesian product of the ranges of these images.

In contrast, for logistic regression, the OR is variation independent of the nuisance model, i.e. log baseline odds: the range of $(\log\{\text{OR}(v)\}, \log\{p_0(v)/(1-p_0(v))\})$ is \mathbb{R}^2 . This may be seen directly in Figure 1 (a) - (c). These L'Abbé plots show $E[Y \mid A = 0]$ on the horizontal vs. $E[Y \mid A = 1]$ on the vertical (L'Abbé et al., 1987). In plots (a) to (c) the vertical lines represent a specific value for the baseline risk $E[Y \mid A=0]$ (or baseline odds); it may be seen that whereas this vertical line intersects every OR curve in Figure 1(c), it clearly does not intersect every RR line in (a), nor every RD line in (b).

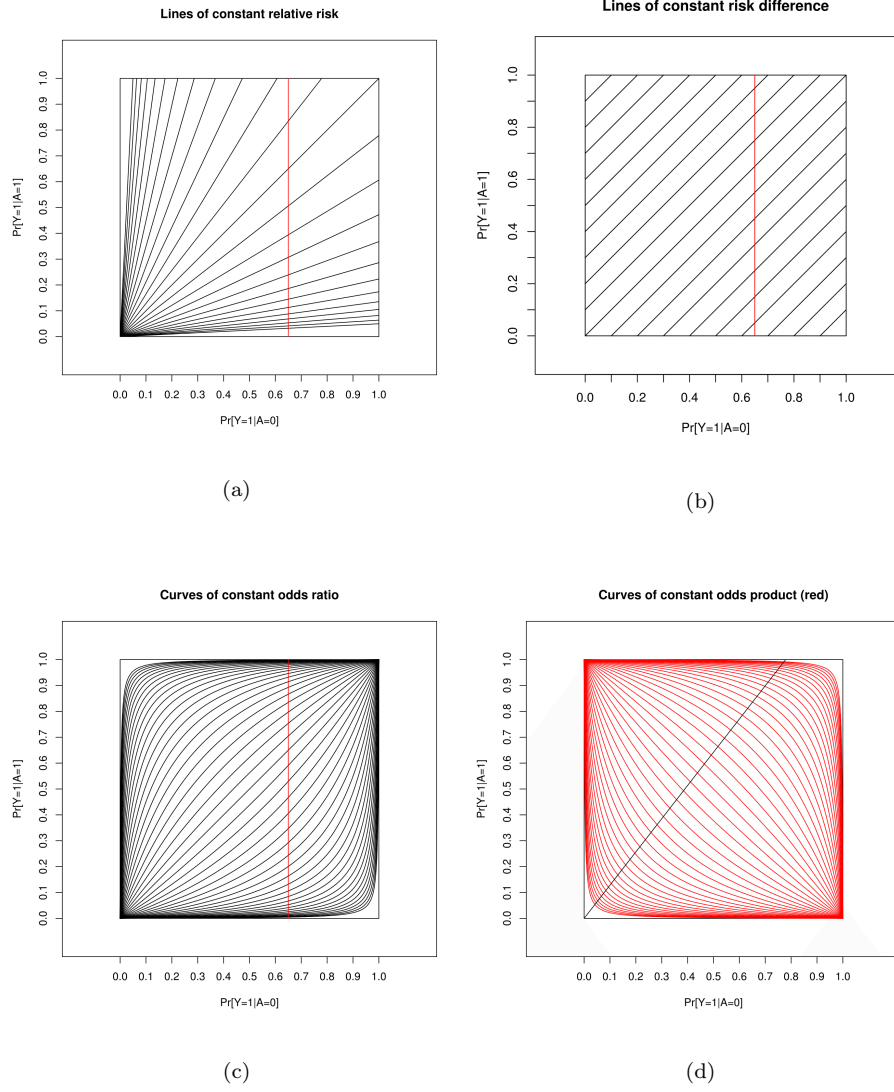


Figure 1: L'Abbé plots: Lines of constant: (a) $\log \text{RR} \in (-3, -2.75, \dots, 3)$ (b) $\text{RD} \in \{-0.9, -0.8, \dots, 0.9\}$. (c) Curves of constant $\log \text{Odds Ratio} \in (-6, -5.75, \dots, 6)$. The vertical lines in these plots represent a baseline risk of 0.65 (or a baseline odds of 1.86). (d) Red curves represent specific values of the Odds Product (OP), $\log \text{OP} \in (-6, -5.75, \dots, 6)$ and the black line passing through the origin in this plot corresponds to a $\log \text{RR}$ of 0.25.

As discussed by many authors, the constraints on the range of $(\log\{\text{RR}(v)\}, \log\{p_0(v)\})$ or $(\text{RD}(v), p_0(v))$ have several undesirable consequences:

- (I) **Models that are always misspecified:** if the support of covariates (W, Z) or parameters (ν, μ) is unbounded, then clearly (1.1) is misspecified *a priori* owing to the variation dependence. Consequently, any *inference* resulting from (1.1) is not reliable.
- (II) **Meaningless predictions:** even if the supports of both the covariates (W, Z) and the parameters (ν, μ) are bounded, the estimated $E[Y|A, V]$ applied to a new sample can predict probabilities that lie outside of the range $[0, 1]$, and hence be nonsensical.

When model (1.1) is used in combination with maximum likelihood for parameter estimation (as is often the case in practice), additional critiques apply:

- (III) **Boundary problems:** the maximum often lies on the boundary of the constrained parameter space. This may give rise to computation difficulties in fitting these models. For example, standard statistical software may report failed convergence when attempting to fit log-binomial models in certain settings (Lumley et al., 2006; McNutt et al., 2003; Williamson et al., 2013).
- (IV) **Lack of robustness to misspecification of nuisance models:** even when the model (1.2) for log RR or RD is correct, misspecification of *the functional form* $\mu^T Z$ for the log baseline risk or baseline risk generally results in inconsistent estimation of the parameter vector ν . This is particularly problematic since, although the log baseline risk or the baseline risk model is required for estimating the RR or RD via maximum likelihood, it is not the target of inference.

We note that although problems (I) - (III) exist with the log-binomial regression/linear probability models these problems do not arise with the logistic and certain other GLMs. However, problem (IV) is shared by MLEs from all GLMs, including logistic regression.

Owing to problems (I) - (III), applied researchers often choose to fit logistic regressions and estimate the OR instead of the RR or RD. Arguably this corresponds to choosing the parameter of interest to make it variation independent of parameters of the simplest

nuisance model, i.e. the baseline risk $p_0(v)$. However, as argued by many authors (e.g., Hellevik, 2009), the choice of statistical model should target the parameter of substantive interest rather than the other way around. Thus our approach is to develop an unconstrained nuisance model $\phi(v)$ that is variation independent of $\text{RR}(v)$ or $\text{RD}(v)$. In this way, we solve problems (I) - (III) without changing the parameter of interest.

On the other hand even with the novel nuisance model, maximum likelihood estimation still suffers from problem (IV). To help alleviate this we propose *doubly-robust* estimators of models for (monotone transforms of) the RR and RD that are consistent and asymptotically normal (CAN) even when the nuisance model $\phi(v)$ is misspecified, provided that we have a correctly specified model for the exposure probability $P(A = 1 | V = v)$, also known as the *propensity score*. In contrast, it is not possible to construct such a doubly-robust estimator for the odds ratio (Tchetgen Tchetgen et al., 2010, Theorem 3). Note that this is of practical importance since in a designed experiment the propensity score will be known by construction.

The rest of this article is organized as follows. In Section 2 we propose our novel nuisance model $\phi(v)$ and briefly outline maximum likelihood estimation. In Section 3, we describe a doubly robust semi-parametric approach to estimation. Sections 4 and 5 contain simulations and an illustrative data analysis. Section 6 contains a summary.

2 Nuisance model specification

To solve problems (I) – (III), the nuisance model $\phi(v)$ needs to be unconstrained and variation independent of (a suitable monotone transformation of) the parameter of interest, $\theta(v)$. Here $\theta(v)$ is a monotone transformation of $\text{RR}(v)$ or $\text{RD}(v)$ that maps the domain of V onto \mathbb{R} . Specifically, our choice for $\theta(v)$ is either $\log\{\text{RR}(v)\}$ or $\text{arctanh}\{\text{RD}(v)\}$ ($= \log\{(1 + \text{RD}(v))/(1 - \text{RD}(v))\}$), and our choice for $\phi(v)$ is the v -specific log Odds Product (OP):

$$\phi(v) \equiv \log \frac{p_0(v)p_1(v)}{(1 - p_0(v))(1 - p_1(v))}.$$

This approach is illustrated in Figure 1(d), where each curve corresponds to a specific value of OP. One can see that $\log(\text{OP})$ ranges from $-\infty$ to ∞ , and each contour line of

OP intersects every RR line (see Fig. 1(a)) and every RD line (see Fig. 1(b)) in exactly one point. Hence if we specify a parametric form for $(\theta(v), \phi(v))$ such that

$$\theta(V) = \alpha^T W, \quad (2.1)$$

$$\phi(V) = \beta^T Z, \quad (2.2)$$

then the parameter of interest, α , and the nuisance parameter, β , are variation independent.

For identification and estimation of the parameters (α, β) , we note that if we choose the log odds product as $\phi(v)$, then the map given by

$$(\theta(v), \phi(v)) \leftrightarrow (P(Y=1|A=0, V=v), P(Y=1|A=1, V=v)), \quad (2.3)$$

is a smooth bijection from \mathbb{R}^2 to $(0, 1) \times (0, 1)$. In fact, the inverse map is given in closed form by solving quadratic equations. Specifically, for $\theta(v) = \log RR(v)$, we have

$$\begin{aligned} p_0(v) &\equiv p(Y=1 | A=0, V=v) \\ &= \frac{-(e^{\theta(v)} + 1)e^{\phi(v)} + \sqrt{e^{2\phi(v)}(e^{\theta(v)} + 1)^2 + 4e^{\theta(v)+\phi(v)}(1 - e^{\phi(v)})}}{2e^{\theta(v)}(1 - e^{\phi(v)})}, \end{aligned} \quad (2.4)$$

$$p_1(v) \equiv p(Y=1 | A=1, V=v) = p_0(v)e^{\theta(v)};$$

similarly for $\theta(v) = \text{arctanh}(RD(v))$ we have

$$\begin{aligned} p_0(v) &\equiv p(Y=1 | A=0, V=v) \\ &= \frac{e^{\phi(v)}(2 - \rho(v)) + \rho(v) - \sqrt{\{e^{\phi(v)}(\rho(v) - 2) - \rho(v)\}^2 + 4e^{\phi(v)}(1 - \rho(v))(1 - e^{\phi(v)})}}{2(e^{\phi(v)} - 1)}, \end{aligned} \quad (2.5)$$

$$p_1(v) \equiv p(Y=1 | A=1, V=v) = p_0(v) + \rho(v),$$

where $\rho(v) \equiv \tanh(\theta(v))$.

As an illustration, Figure 2 plots $p_1(v)$ as functions of $\theta(v)$ and $\phi(v)$, where $\theta(v) = \log RR(v)$. Plots with $p_0(v)$ are similar in shapes and hence omitted, and plots with risk difference are given in Appendix A. Under the proposed models, the inverse maps (2.4) and (2.5) are sigmoid functions of $\theta(v)$ and $\phi(v)$; these are similar to the case with a logistic

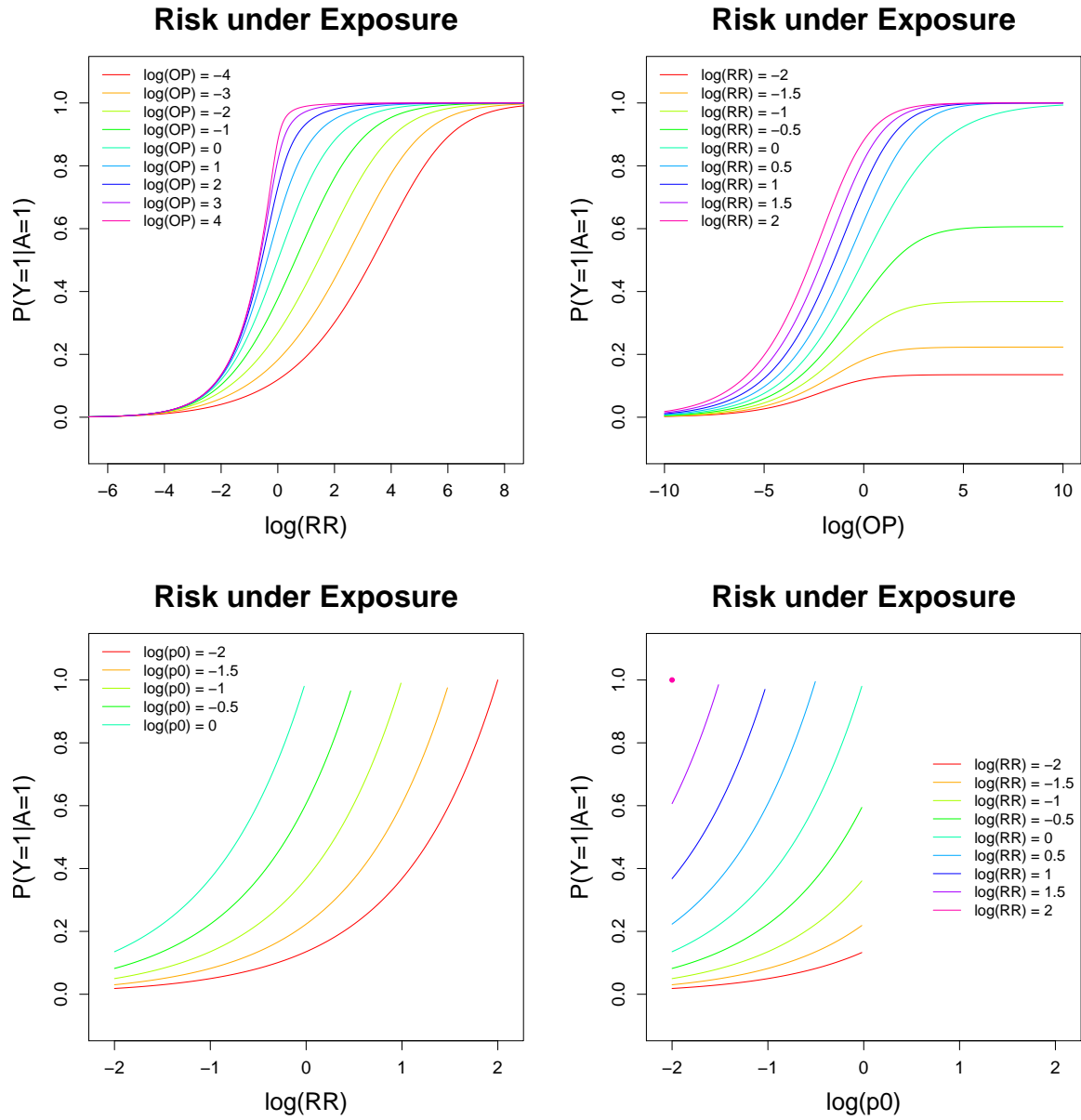


Figure 2: Fitted probabilities from the proposed relative risk model (upper panels) and the Poisson mean model (lower panels). For the proposed relative risk model, curves for fitted risks under non-exposure have the same shapes to those for fitted risk under exposure.

regression model. In contrast, as one can see from the lower panels of Figure 2, the inverse maps from a Poisson mean model do not level off even when the probabilities approach 1.

Under the model given by (2.1) and (2.2), the parameters α and β may be estimated directly via maximum likelihood without the need for constrained maximization. The log-likelihood for one observation is given by:

$$\ell(\alpha, \beta) = Y \log(P(Y=1|A, V; \alpha, \beta)) + (1 - Y) \log(P(Y=0|A, V; \alpha, \beta)).$$

Likelihood-based confidence intervals for α and β can then be obtained in standard fashion.

We close this section with several remarks. First, there are other nuisance parameters that are also unconstrained and variation independent of the (log) relative risk. For example Tchetgen Tchetgen (2013) considers an alternative nuisance parameter, $\log(P(Y = 1|A = 0, V)/P(Y = 1|A = 0))$, for estimating the relative risk. Similar to the log odds product, his nuisance parameter is also unconstrained and variation independent of the (log) relative risk. However, his nuisance parameter does not give rise to the full likelihood. In other words, with his nuisance parameter, equation (2.3) is no longer a bijection from \mathbb{R}^2 to $(0, 1) \times (0, 1)$. Consequently, one cannot make predictions or use maximum likelihood for parameter estimation.

There are clearly many choices for $\phi(v)$ that would make the bivariate map (2.3) a smooth bijection. However, the log odds product has the advantage that it is simple and the inverse map can be obtained in a closed form. It is worth noting that the odds product is used solely as the nuisance model to form a parameterization in conjunction with (a suitable monotone transformation of) the RR or RD, which is the target of inference. Thus the fact that the odds product is nonlinear and hence not collapsible does not lead to problems of interpretation for $RR(v)$ or $RD(v)$. Furthermore, formulating a model for the log odds product is similar in spirit to the specification of a logistic regression model. In fact, suppose one specifies a logistic model for the variable $\tilde{Y} \triangleq AY + (1 - A)(1 - Y)$, such as $\text{logit}(P(\tilde{Y} = 1|A, V)) = \tilde{\alpha}^T W + \beta^T AZ$, then this implies a model for the log odds product (of outcome Y): $\log OP(V) = \beta^T Z$. Note here \tilde{Y} is an indicator function that the observed outcome Y “agrees” with the treatment indicator A . Consideration of \tilde{Y} here is somewhat similar in spirit to the ‘switch relative risk’ (Van Der Laan et al., 2007).

Finally in the next section we address problem (IV) by introducing a doubly robust

estimating procedure that provides an extra layer of protection against misspecification of the log odds product model.

3 Semiparametric estimation

In the remainder of the paper we will always take model (2.1) to be true, so that α is well-defined. If the nuisance model (2.2) is misspecified, the maximum likelihood estimate for α may not be consistent. To address this we propose an estimator that is consistent for α , even if (2.2) is incorrect, provided that we have correctly specified a model for the *propensity score*:

$$e(V) \equiv P(A = 1 | V). \quad (3.1)$$

In more detail, we construct a locally semiparametric efficient estimator and show that it is CAN if either (but not necessarily both) model (2.2) *or* (3.1) is correct. Such estimators are commonly referred to as *doubly robust* (Robins and Rotnitzky, 2001; Van der Laan and Robins, 2003).

The following theorem provides the basis for our estimator. Throughout this section, we make the positivity assumption that with probability 1, $e(V), p_0(V), p_1(V) \in (0, 1)$.

Theorem 3.1 *Let $\dot{\alpha}$ be the unknown true value of α , and $\dot{\eta}$ be the unknown true value of an infinite dimensional nuisance parameter consisting of $\phi(\cdot), e(\cdot)$ and the marginal distribution of V . In the semiparametric model characterized by the sole restriction (2.1), the efficient score for α under the law $(\dot{\alpha}, \dot{\eta})$ is:*

$$S^*(\dot{\alpha}; \dot{\eta}) = \omega_{\text{eff}}(V)(A - e(V))(H(\dot{\alpha}) - p_0(V)),$$

where $\omega_{\text{eff}}(V)$ is the efficient weighting function given in Appendix B; $H(\alpha) = Y \{\exp(-A\alpha^T W)\}$ when $\theta(V) = \log(\text{RR}(V))$, and $H(\alpha) = Y - A \tanh(\alpha^T W)$ when $\theta(V) = \text{arctanh}(\text{RD}(V))$.

Further, if for some $\sigma > 0$,

$$e(V), p_0(V) \in (\sigma, 1 - \sigma) \quad (3.2)$$

with probability 1, then the efficient score is bounded.

The proof is left to Appendix B. It follows that the solution $\hat{\alpha}_{\text{eff}}$ to

$$\mathbb{P}_n [\omega_{\text{eff}}(V)(A - e(V))(H(\alpha) - p_0(V))] = 0,$$

has asymptotic variance equal to the semiparametric variance bound in the model defined by the sole restriction (2.1); here $\mathbb{P}_n(\cdot)$ denotes the empirical expectation.

Of course $\hat{\alpha}_{eff}$ is not feasible as it depends on unknown quantities $e(V)$ and $p_0(V)$. When V is high dimensional or has several continuous elements, common practice is to specify a parametric model for these population quantities. For example, one may specify models (2.1) and (2.2) and estimate $p_0(V)$ from (2.4) and (2.5) for the RR and RD, respectively. One may also specify a parametric model for the propensity score $e(V)$, such as

$$e(V; \gamma) \equiv P(A = 1|V; \gamma) = \text{expit}(\gamma^T X), \quad (3.3)$$

where $X = x(V)$ is a known vector function of V and γ is a finite dimensional parameter.

Let $\hat{\gamma}$ solve

$$\mathbb{P}_n[X\{A - e(V; \gamma)\}] = 0,$$

and $(\hat{\alpha}, \hat{\beta})$ solve

$$0 = \mathbb{P}_n \left[\frac{\partial}{\partial(\alpha, \beta)} \ell(\alpha, \beta) \right] \equiv \mathbb{P}_n[S(\alpha, \beta)].$$

Note that here $\hat{\alpha}$ is the maximum likelihood estimator of Section 2. Under suitable regularity conditions (White, 1982) the estimators $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ converge in probability to fixed constants $\alpha^*, \beta^*, \gamma^*$ regardless of whether the models (2.2) or (3.3) are correct or not. However, α^* may not be equal to the true value of α if the nuisance model (2.2) is misspecified.

Since $\hat{\alpha}$ may be inconsistent, we instead estimate α by the solution $\hat{\alpha}_{DR, eff}$ to the following estimating equation:

$$\mathbb{P}_n[U(\alpha; \hat{\alpha}, \hat{\beta}, \hat{\gamma})] = 0,$$

where

$$U(\alpha; \hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \omega_{eff}(V; \hat{\alpha}, \hat{\beta}, \hat{\gamma}) (A - e(V; \hat{\gamma})) \left(H(\alpha) - p_0(V; \hat{\alpha}, \hat{\beta}) \right). \quad (3.4)$$

Under correct specification of model (2.1), $\hat{\alpha}_{DR, eff}$ is consistent if either the nuisance model (2.2) or the propensity score model (3.3) is correct. Furthermore, if we replace the efficient weighting function in (3.4) with any other function $\omega(V)$, under regularity conditions the solution $\hat{\alpha}_{DR}(\omega)$ to

$$\mathbb{P}_n[U(\alpha; \hat{\alpha}, \hat{\beta}, \hat{\gamma}; \omega)] = 0, \quad (3.5)$$

where

$$U(\alpha; \hat{\alpha}, \hat{\beta}, \hat{\gamma}; \omega) = \omega(V) (A - e(V; \hat{\gamma})) \left(H(\alpha) - p_0(V; \hat{\alpha}, \hat{\beta}) \right) \quad (3.6)$$

also yields a doubly robust estimator. These results are formally described in the following theorem. The proof is left to Appendix C. We define the *union model* to be the set of distributions characterized by (2.1) and *either* model (2.2) *or* the propensity model (3.3) being correct.

Theorem 3.2 *Under standard regularity conditions and the positivity assumption (3.2), $\dot{\alpha}$ is the unique solution to the probability limit of (3.5) in the union model. If $\hat{\alpha}_{DR}(\omega)$ solves (3.5) then $\hat{\alpha}_{DR}(\omega)$ is a regular and asymptotically linear (RAL) estimator of α under this model.*

Furthermore, the influence function of $\hat{\alpha}_{DR}(\omega)$ is given by

$$\tau^{-1} \tilde{U}(\dot{\alpha}; \alpha^*, \beta^*, \gamma^*; \omega), \quad (3.7)$$

where

$$\tau = -E [\partial U(\alpha; \alpha^*, \beta^*, \gamma^*; \omega) / \partial \alpha |_{\alpha=\dot{\alpha}}]$$

and

$$\begin{aligned} \tilde{U}(\dot{\alpha}; \alpha^*, \beta^*, \gamma^*; \omega) &= U(\dot{\alpha}; \alpha^*, \beta^*, \gamma^*; \omega) - \\ &E \left[\frac{\partial U(\dot{\alpha}; \alpha, \beta, \gamma^*; \omega)}{\partial(\alpha^T, \beta^T)} \middle| \begin{matrix} \alpha = \alpha^* \\ \beta = \beta^* \end{matrix} \right] \times \left\{ E \left[\frac{\partial S(\alpha, \beta)}{\partial(\alpha^T, \beta^T)} \middle| \begin{matrix} \alpha = \alpha^* \\ \beta = \beta^* \end{matrix} \right] \right\}^{-1} S(\alpha^*, \beta^*) - \\ &E \left[\frac{\partial U(\dot{\alpha}; \alpha^*, \beta^*, \gamma; \omega)}{\partial \gamma} \middle|_{\gamma=\gamma^*} \right] \times \left\{ E \left[-X \frac{\partial e(V; \gamma)}{\partial \gamma} \middle|_{\gamma=\gamma^*} \right] \right\}^{-1} X \{A - e(V; \gamma^*)\}. \end{aligned}$$

The estimator $\hat{\alpha}_{DR}(\omega)$ is only one of many estimators using the weight function ω that are doubly-robust (i.e. regular in the union model); $\hat{\alpha}_{DR}(\omega)$ is arguably the simplest, both conceptually and computationally. In the last decade there has been an explosion of

research that has, in a number of semi-parametric models, developed a methodology for constructing doubly robust estimators that have much better large sample efficiency (at laws outside of the intersection sub-model) and improved finite sample performance compared to the simplest DR estimators; Rotnitzky and Vansteelandt (2014) give a review. Though these methods could also be applied in our context, we leave this to future work as it is tangential to our goal of demonstrating the utility of our novel nuisance modeling.

Remark 3.3 We note that double robustness is a useful property only if it is possible for either nuisance model to be correct *a priori*. In other words, for the double robustness property to be relevant, equation (3.5) must be used in combination with an (implicit) model for baseline risk that does not suffer from problem (I). Our approach based on the log OP avoids problem (I) due to the variation independence between parameters in model (2.1) and (2.2).

Remark 3.4 One can show that the efficient influence function in the union model is also $S^*(\dot{\alpha}; \dot{\eta})$ at the submodel where the propensity score (3.3) is correctly specified. Consequently, the semiparametric variance bound is the same for the union model and the model defined by the sole restriction (2.1) when the propensity score model is correct. In particular, they are the same at the intersection model where both (2.2) and (3.3) are correctly specified. See Robins et al. (1994) for related results in a general context.

Regardless of whether or not the union model is correct, the influence function of $\hat{\alpha}_{DR}(\omega)$ for fixed ω is given by (3.7), replacing $\dot{\alpha}$ by the probability limit of $\hat{\alpha}_{DR}(\omega)$, assuming that this exists. Hence even under mis-specification of both nuisance models a consistent estimator of the asymptotic variance of $n^{1/2}(\hat{\alpha}_{DR}(\omega) - \dot{\alpha})$ is still $\hat{\tau}^{-1}\hat{\Sigma}\hat{\tau}^{-1}$ with

$$\hat{\Sigma} = \mathbb{P}_n \left[\hat{\tilde{U}}(\hat{\alpha}_{DR}(\omega); \hat{\alpha}, \hat{\beta}, \hat{\gamma}; \omega) \hat{\tilde{U}}(\hat{\alpha}_{DR}(\omega); \hat{\alpha}, \hat{\beta}, \hat{\gamma}; \omega)^T \right],$$

where $\hat{\tilde{U}}$ is \tilde{U} with all expectations replaced by sample averages and $(\dot{\alpha}, \alpha^*, \beta^*, \gamma^*)$ replaced by $(\hat{\alpha}_{DR}(\omega), \hat{\alpha}, \hat{\beta}, \hat{\gamma})$, and

$$\hat{\tau} = -\mathbb{P}_n \left[\partial U(\alpha; \hat{\alpha}, \hat{\beta}, \hat{\gamma}; \omega) / \partial \alpha^T \Big|_{\alpha = \hat{\alpha}_{DR}(\omega)} \right].$$

Remark 3.5 The influence function of $\hat{\alpha}_{DR,eff}$ can be obtained in two steps: 1) replace ω in each term of (3.7), τ and $\tilde{U}(\dot{\alpha}; \alpha^*, \beta^*, \gamma^*; \omega)$ with $\omega_{eff}(\alpha, \beta, \gamma)$; 2) evaluate (3.7) at the point where $(\alpha, \beta, \gamma) = (\alpha^*, \beta^*, \gamma^*)$. A consistent estimator of the asymptotic variance of $n^{1/2}(\hat{\alpha}_{DR,eff} - \dot{\alpha})$ can then be obtained from the influence function analogously to that for $\hat{\alpha}_{DR}(\omega)$. Note as ω_{eff} depends on parameters α, β, γ , the derivatives following equation (3.7) should in general include the derivatives of ω_{eff} with respect to these parameters. However when the union model is true, these derivatives are not required as they do not contribute to the asymptotic variance.

The following theorem states that $\hat{\alpha}_{DR,eff}$ is locally semiparametric efficient in the union model at the intersection submodel. The proof is left to Appendix D.

Theorem 3.6 *If both the nuisance model (2.2) and the propensity score model (3.3) are correct, then the variance of $\hat{\alpha}_{DR,eff}$ attains the semiparametric variance bound under the union model.*

Remark 3.7 The asymptotic variance of $\hat{\alpha}_{DR,eff}$ based on a *correct* log OP model and an *incorrect* propensity score model may be smaller than the asymptotic variance (under the same law) with both models correctly specified. On the other hand, the asymptotic variance of $\hat{\alpha}_{DR,eff}$ based on an *incorrect* log OP model and a *correct* propensity score model is never smaller than that obtained under correct specification of both models. See Robins and Rotnitzky (2001) for related results in a general context.

4 Simulation studies

In this section, we evaluate the finite sample performance of our proposed procedure. We generate data from the models (2.1), (2.2) and (3.3), where $\dot{\alpha} = (0, -1)^T$, $\dot{\beta} = (-0.5, 1)^T$ and $\dot{\gamma} = (0.1, -0.5)^T$. The covariates V include an intercept and a uniform random variable generated from $\text{Unif}(-2, 2)$; $w(V)$, $z(V)$ and $x(V)$ are identity functions of V .

We also consider scenarios in which the nuisance models, (2.2) and (3.3) are misspecified. In particular, instead of using V , the analyst uses covariates V^\dagger , which include an intercept and an irrelevant covariate also generated from $\text{Unif}(-2, 2)$; the misspecified nuisance models (2.2) and (3.3) are, respectively, linear and logistic in V^\dagger .

We consider three estimators:

mle: The maximum likelihood estimator;

drw: The optimally weighted doubly-robust estimator;

dru: The doubly-robust estimator $\hat{\alpha}_{DR}(\omega)$ with the (naive) weighting function given by $\omega(V) = V$.

The standard deviation of **mle** is estimated via inverting the Fisher information matrix, and the standard deviations of **drw** and **dru** are estimated using the methods described in and right before Remark 3.5.

We consider four scenarios:

bth: V is used in both the nuisance model (2.2) and the propensity score model (3.3);

psc: V is used in the propensity score model (3.3) but V^\dagger is used in the nuisance model (2.2);

orc: V is used in the nuisance model (2.2) but V^\dagger is used in the propensity score model (3.3);

bad: V^\dagger is used in both the nuisance model (2.2) and the propensity score model (3.3).

Note that because the propensity score is not used in **mle**, results for **mle.orc** are identical to **mle.bth**, and results for **mle.psc** are identical to **mle.bad**. In our simulations we suppose that the model for the target of interest, (2.1), is correctly specified. All the simulation results are based on 1000 Monte-Carlo runs of $n = 500$ and $n = 1000$ units each.

Table 1 summarizes simulation results. All estimators and scenarios have bias smaller than 0.03, except for **mle.bad** and **drw.bad**, the latter confirms “double robustness.” With both nuisance models correctly specified, the optimally weighted estimator **drw.bth** has standard deviation smaller than, or comparable to, **dru.bth** for both the RR and RD at both sample sizes, showing that there is indeed an efficiency gain achieved by using the optimal weighting function. In our simulations, results for **dru** under misspecification were very similar to **drw** and hence are omitted. We also note that although α_0 and α_1 take the same true values in the relative risk and risk difference model, they have different

Table 1: Monte Carlo results of the proposed MLE and DR estimators. The true values for α_0 and α_1 are 0 and -1, respectively for both RR and RD. The sample size is 500.

	Relative Risk		Risk Difference	
	α_0	α_1	α_0	α_1
Bias(SE)				
mle.bth	0.008(0.004)	-0.021(0.005)	0.004(0.002)	-0.009(0.003)
mle.bad	-0.403(0.004)	0.021(0.004)	-0.028(0.002)	-0.007(0.003)
drw.bth	0.008(0.004)	-0.023(0.005)	0.004(0.002)	-0.009(0.003)
drw.psc	0.007(0.004)	-0.022(0.005)	0.003(0.002)	-0.010(0.003)
drw.lop	0.008(0.004)	-0.024(0.005)	0.004(0.002)	-0.009(0.003)
drw.bad	-0.101(0.004)	0.024(0.005)	-0.029(0.002)	-0.016(0.003)
dru.bth	0.019(0.005)	-0.057(0.008)	0.004(0.002)	-0.010(0.003)
SD Accuracy*				
mle.bth	0.991	0.968	1.015	1.017
mle.bad	0.984	0.949	1.034	1.015
drw.bth	0.996	0.971	1.015	1.007
drw.psc	1.000	0.986	1.011	0.998
drw.orc	0.991	0.954	1.017	1.010
drw.bad	0.946	0.953	1.023	1.010
dru.bth	0.989	0.808	1.014	0.999
Coverage**				
mle.bth	0.955	0.953	0.963	0.952
mle.bad	0.043	0.928	0.938	0.951
drw.bth	0.958	0.956	0.960	0.940
drw.psc	0.961	0.963	0.960	0.944
drw.orc	0.960	0.946	0.960	0.953
drw.bad	0.845	0.935	0.938	0.952
dru.bth	0.958	0.949	0.959	0.944

*: SD Accuracy = Estimated SD / Monte Carlo SD.

**: Nominal level = 95%.

interpretations. Hence it is not (practically) relevant to directly compare the simulation results across different target parameters (i.e. RR and RD).

Theory predicts that under correct specification of both the regression model and the propensity score model (**bth**), **mle** achieves the efficiency bound for the parametric outcome regression model, while **drw** achieves the efficiency bound in the larger union model. This is consistent with our simulation results: the standard deviation of **mle.bth** is no larger than that of **drw.bth**.

We then evaluate the accuracy of proposed standard deviation estimator, defined as the ratio between the estimated standard deviation and the Monte Carlo standard deviation. Simulation results show that at modest sample size, the estimated standard deviation generally provides a good approximation to the Monte Carlo standard deviation, especially when at least one of models (2.2) and (3.3) is correctly specified. Although the standard deviation estimator associated with **dru.bth** is slightly biased downwards for estimating α_1 in the relative risk model at sample size 500, nonetheless the bias is not sufficient to distort the coverage properties of our interval estimators and it decreases with sample size. For example, the accuracy of the proposed standard deviation estimator is 0.965 if the sample size is 1000, and 0.980 if the sample size is 10,000. These results are consistent with our claims right after Remark 3.4. We also compute nominal 95% Wald-type confidence intervals based on point and standard deviation estimates. Nominal coverage rates are achieved for all scenarios except for **mle.bad** and **drw.bad**.

To investigate the sensitivity of the proposed estimator to misspecification of the *functional form* of the nuisance model, we further consider a simulation scenario where the data are generated from models (2.1), (3.3) and a baseline model that $p_0(V) = \gamma^T z(V)$, where $\dot{\alpha} = (0, 0.3)^T$, $\dot{\beta} = (0.5, 0.2)^T$ and $\dot{\gamma} = (0.1, -0.5)$. The covariates V include an intercept and a uniform random variable generated from $\text{Unif}(-1, 1)$; $w(V)$, $z(V)$ and $x(V)$ are identity functions of V . These coefficient values and covariate ranges are carefully chosen to ensure that the underlying true values of $p_1(V)$ and $p_0(V)$ fall into the unit interval for all possible values of V . For comparison, we also fit the true model according to the data generating mechanism; that is, we replace the log odds product model in the proposed approach with a linear baseline risk model.

Table 2: Monte Carlo bias and standard error (in parenthesis) with different nuisance models. The data is generated following a linear nuisance model on p_0 , whereas the proposed approach uses a linear nuisance model on $\log(OP)$. The true values for α_0 and α_1 are 0 and 0.3, respectively for both RR and RD. The sample size is 500.

Nuisance		Relative Risk		Risk Difference	
		α_0	α_1	α_0	α_1
mle.bth	$\log(OP)$	-0.003(0.003)	-0.004(0.005)	-0.002(0.001)	-0.001(0.002)
mle.bth	p_0	-0.003(0.003)	-0.003(0.005)	-0.001(0.001)	-0.010(0.002)
mle.bad	$\log(OP)$	-0.042(0.003)	0.044(0.005)	-0.062(0.001)	0.029(0.003)
mle.bad	p_0	-0.073(0.003)	0.357(0.003)	-0.019(0.001)	0.193(0.001)
drw.bth	$\log(OP)$	-0.004(0.003)	0.002(0.005)	-0.002(0.001)	0.000(0.002)
drw.bth	p_0	-0.004(0.003)	0.003(0.005)	-0.002(0.001)	0.000(0.002)
drw.psc	$\log(OP)$	-0.004(0.003)	0.000(0.005)	-0.002(0.001)	-0.001(0.002)
drw.psc	p_0	-0.008(0.003)	0.000(0.005)	-0.002(0.001)	-0.003(0.002)
drw.orc	$\log(OP)$	0.000(0.003)	-0.015(0.005)	-0.002(0.001)	-0.001(0.002)
drw.orc	p_0	-0.003(0.003)	0.002(0.005)	-0.002(0.001)	-0.001(0.002)
drw.bad	$\log(OP)$	-0.097(0.003)	0.027(0.005)	-0.061(0.001)	-0.007(0.002)
drw.bad	p_0	-0.071(0.003)	-0.014(0.005)	-0.040(0.001)	-0.009(0.002)

Table 2 summarizes the simulation results. Similar to the previous simulation, we consider four scenarios: **bth**, **psc**, **orc**, **bad**, with the only difference being that the nuisance model can now be either the linear log odds product model or the linear baseline risk model; the latter model is a reasonable alternative here owing to the restricted ranges of the covariates. All findings are compatible with our theoretical development. We note further that in the context of our simulation, when V^\dagger is used in both the nuisance model and the propensity score model (i.e. **bad**), our proposed nuisance model is more robust with the maximum likelihood estimator, whereas a linear nuisance model on the baseline risk appears to be more robust with the doubly robust estimator. Also with the proposed $\log(OP)$ nuisance model, the estimation results are moderately biased for **drw.orc**. This is because contrary to the data-generating process, the $\log(OP)$ model implies a non-linear model (2.4) or (2.5) on the baseline risk. Thus even if we include covariate V in the linear log odds product model, the nuisance model is still misspecified in its functional form, which leads to bias in estimation. Finally, the bias of **drw.orc** with $\log(OP)$ is smaller compared to the bias of **drw.bad** with p_0 , suggesting that under our simulation setting, bias from omitting a relevant variable is larger than bias from misspecification of the functional form of the nuisance model.

5 Application to fetal data

We illustrate the proposed novel models with data from an obstetric study. The data set consists of observations on 14,484 women who delivered at Beth Israel Hospital, Boston from Jan 1970 to Dec 1975. Of these women, 50.4% received electronic fetal monitoring (EFM). The scientific goal is to evaluate the impact of EFM on cesarean section (CS) rates, which is quantified by the ratio of CS rates for monitored and unmonitored women. It was found previously that to avoid potential confounding bias, it was sufficient to control for 4 variables: nulliparity (**nulli**), arrest of labor progression (**arrest**), breech (**breech**) and year of study (**year**) (Neutra et al., 1980). As there was extreme heterogeneity across groups defined by these variables, we let the covariates W and Z be identical and include the main terms of these four variables as well as all two-way and three-way interaction terms of **nulli**, **arrest** and **breech**. The variable **year** was included in the model as a continuous variable.

We refer interested readers to Neutra et al. (1980) for more details of the study.

In our analysis, we applied the following seven methods to the fetal data set: i) **mle**: the proposed RR model fitted using MLE; ii) **dr**: the proposed RR model fitted using DR estimation with the optimal weights; iii) **dr.un**: the proposed RR model fitted using DR estimation, with weight matrix given by the covariates; iv) **poisson**: Poisson regression model fitted using MLE and robust sandwich standard error estimates; v) **log-binomial**: log-binomial regression model fitted using MLE; vi) **dr.p0** DR estimation with a linear nuisance model on the baseline risk, with weight matrix given by the covariates; vii) **logistic**: fit a logistic regression model; then estimate the conditional RR for a given subgroup from the fitted risks. Models iv), v), vii) are fitted with the **glm** function in R.

Figure 3 summarizes the fitted risks from the proposed RR model in different subgroups. The parameters were estimated using maximum likelihood estimation. As expected, the CS rates were very low among women with no complication, but were very high among nulliparas with arrest of labor progression and breech presentation. The effect of EFM was very different across confounder groups. EFM was associated with a decrease in CS rates in the **breech** only, **nulli+breech** and **nulli+arrest+breech** group, but was associated with an increase in CS rates in the **arrest** only and **arrest+breech** group. These findings are consistent with the opposing effects reported in Neutra et al. (1980).

Table 3 summarizes the comparison between the proposed RR model and popular GLM models. The bootstrapped estimates are based on 1000 nonparametric bootstrap samples. The log-binomial regression model failed to give an answer as the program could not work out reasonable starting values. The logistic regression model could not be directly used to estimate the dependence of RR on the covariates. The rest of the models gave comparable point estimates. For the proposed estimators, the bootstrapped standard error estimates agree with the proposed standard error estimates except for interaction terms **br*arr** and **arr*br*nul**. Note that only 83 women had both a breech presentation and arrest of labor progression; within them, only 54 women were nulliparas. Hence this discrepancy in standard error estimates is likely due to the small sample size for these subgroups. In contrast, for the Poisson model, the bootstrapped standard errors are substantially smaller than the robust sandwich error estimates for almost all covariates. Furthermore, for inter-

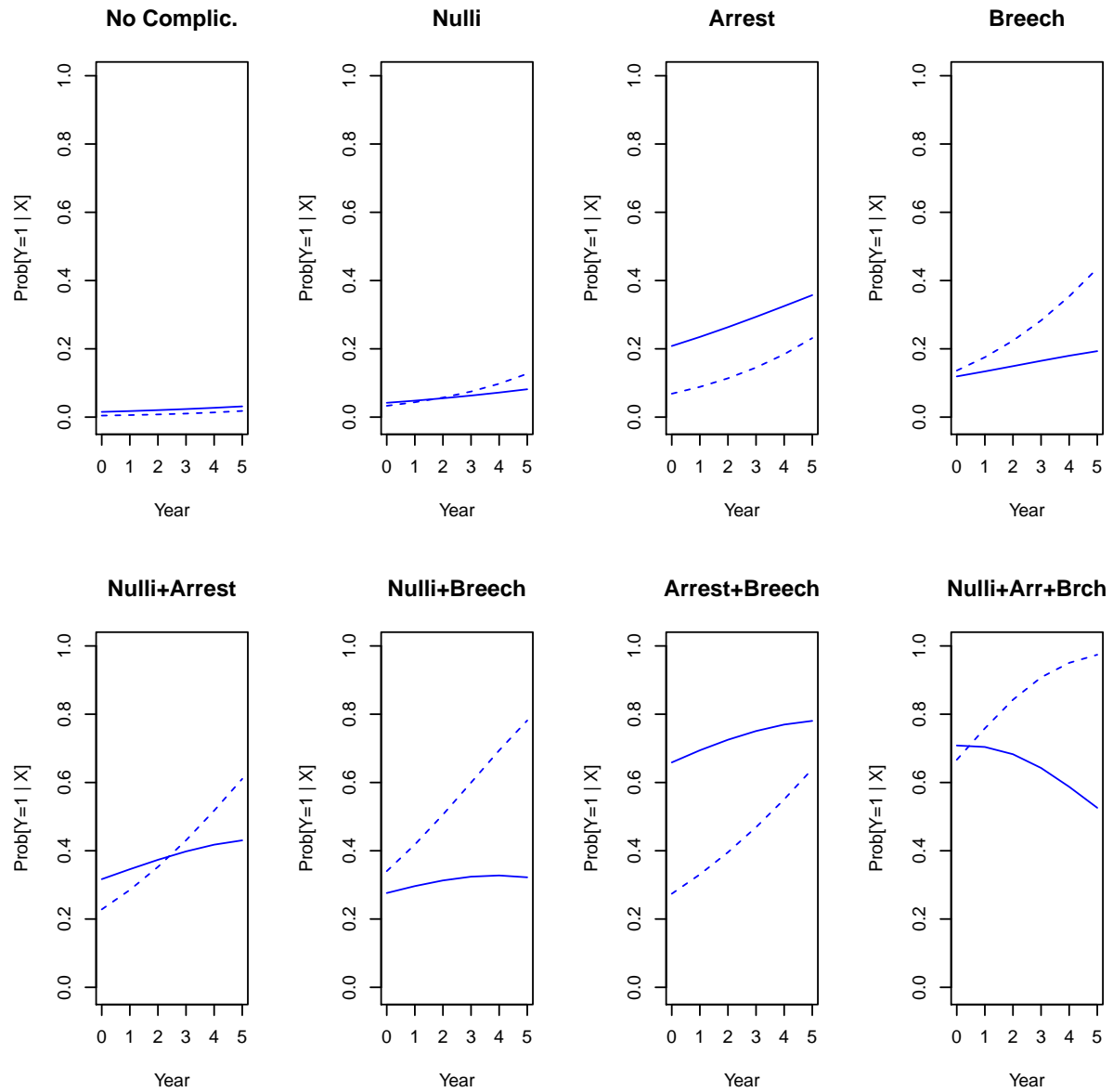


Figure 3: Estimated risks of cesarean section rates by subgroups. The dashed lines denote the unmonitored groups, while the solid lines denote the monitored groups. The model parameters are fitted with maximum likelihood.

Table 3: Comparison of model coefficients estimated using different methods

	intercept	year	nulli	arrest	breech	br*arr	arr*nulli	br*nulli	arr*br*nul
Point estimate									
mle	1.234	-0.136	-0.999	-0.122	-1.368	1.133	0.215	0.924	-0.957
dr	1.339	-0.106	-1.162	-0.297	-1.551	1.109	0.347	1.005	-0.833
dr.un	1.375	-0.110	-1.190	-0.351	-1.584	1.170	0.410	1.047	-0.933
poisson	1.291	-0.140	-1.002	-0.118	-1.393	0.960	0.148	0.892	-0.781
dr.p0	1.463	-0.112	-1.277	-0.440	-1.694	1.334	0.487	1.153	-1.090
Estimated SE*									
mle	0.229	0.029	0.239	0.344	0.340	0.597	0.371	0.396	0.656
dr	0.232	0.036	0.244	0.360	0.351	0.586	0.388	0.404	0.647
dr.un	0.243	0.037	0.251	0.356	0.351	0.583	0.385	0.406	0.646
poisson	0.912	0.167	1.069	1.024	1.008	1.288	1.359	1.332	1.660
Bootstrapped SE									
mle	0.235	0.030	0.236	0.370	0.348	0.833	0.392	0.404	0.881
dr	0.235	0.036	0.240	0.385	0.361	0.895	0.406	0.415	0.940
dr.un	0.245	0.037	0.247	0.378	0.361	0.985	0.401	0.418	1.034
poisson	0.234	0.029	0.237	0.373	0.349	1.744	0.396	0.405	1.778
dr.p0	0.270	0.038	0.271	0.394	0.378	1.994	0.417	0.438	2.028

*: As there is no readily applicable software for computing the standard error of coefficients estimated with **dr.p0**, we only report the bootstrapped estimates for the standard errors.

action terms **br*arr** and **arr*br*nul**, the bootstrapped standard errors of Poisson regression estimates are much more larger than those of the proposed RR model. Similarly with the **dr.p0** method. These suggest that estimates with the proposed RR model may be more stable in small sample settings.

We further illustrate these methods by examining a particular subgroup. As an example, we estimated the relative risk associated with EFM for women who visited the hospital in 1970 and had no complications. In other words, these women are not nulliparas and had normal presentation and labor progression. Bootstrapping method is used for estimating the standard errors and confidence intervals. One can see from Table 4 that all five methods

Table 4: Comparison of inference results on women delivered in 1970 with no complications

	Relative risk estimate	95% Confidence interval	P-value
mle	3.434	[2.164,5.447]	< .001
dr	3.815	[2.405,6.051]	< .001
dr.un	3.953	[2.446,6.391]	< .001
poisson	3.635	[2.298,5.749]	< .001
dr.p0	4.321	[2.545,7.335]	< .001
logistic	4.153	[1.792,6.515]	0.009

suggest that for this particular subgroup, the CS rate in the monitored group is estimated to be around 3.43 to 4.32 folds of that in the unmonitored group. However, the 95% confidence intervals with the `dr.p0` and `logistic` methods are wider than the others.

6 Discussion

We have presented a general approach to modeling the RR and RD as functions of baseline covariates. Our results fill an important methodological gap since many analysts report odds ratios as it is simple to model the dependence of the OR on baseline covariates via logistic regression even though many practitioners regard the RR and RD as much more interpretable than odds ratios. We have also described methods that are consistent for estimating the RR and RD under correct specification of models for these quantities in conjunction with either a correct model for the propensity score or the odds product. The code for implementing our methods is available in a new R package `brm` (short for “binomial regression models”).

The key methodological insight in this paper is that a non-GLM approach may be used to solve the dilemma between choosing a parameter that is collapsible while wishing to have a model with an unconstrained parameter space. By avoiding the GLM we are able to separate the target model from the nuisance model.

In this paper, we have assumed that the observed data is a representative sample from the population of interest. The proposed models can however, be extended to estimate the

RR and RD in case-control studies in conjunction with a model for the sampling probability even without a rare disease assumption.

Appendix

A Plots for fitted probabilities from the proposed risk difference models

Figure 4 plots $p_1(v)$ as functions of $\theta(v)$ and $\phi(v)$, where $\theta(v) = \operatorname{arctanh} RD(v)$. Plots with $p_0(v)$ are similar in shapes and hence omitted. Under the proposed models, the inverse maps (2.4) and (2.5) are sigmoid functions of $\theta(v)$ and $\phi(v)$; these are similar to the case with a logistic regression model. In contrast, as one can see from the lower panels of Figure 4, the inverse maps from a linear mean model do not level off even when the probabilities approach 1.

B Proof of Theorem 3.1

Proof.

Efficient score for conditional relative risk

Van der Laan and Rose (2011, §A.15) showed that the efficient score for the conditional relative risk is

$$S^*(\dot{\alpha}, \dot{\eta}) = h^*(A|V) (H(\dot{\alpha}) - p_0(V)),$$

where

$$h^*(A|V) = \frac{p_A}{p_0(1-p_A)} \left(AW - \frac{E \left[\frac{AW p_A}{1-p_A} \mid V \right]}{E \left[\frac{p_A}{1-p_A} \mid V \right]} \right),$$

$p_0 \equiv p_0(V)$ and $p_A \equiv P(Y = 1|A, V)$. By analytic calculation, it is straightforward to verify that

$$h^*(A|V) = \omega_{eff}(V)(A - e(V)),$$

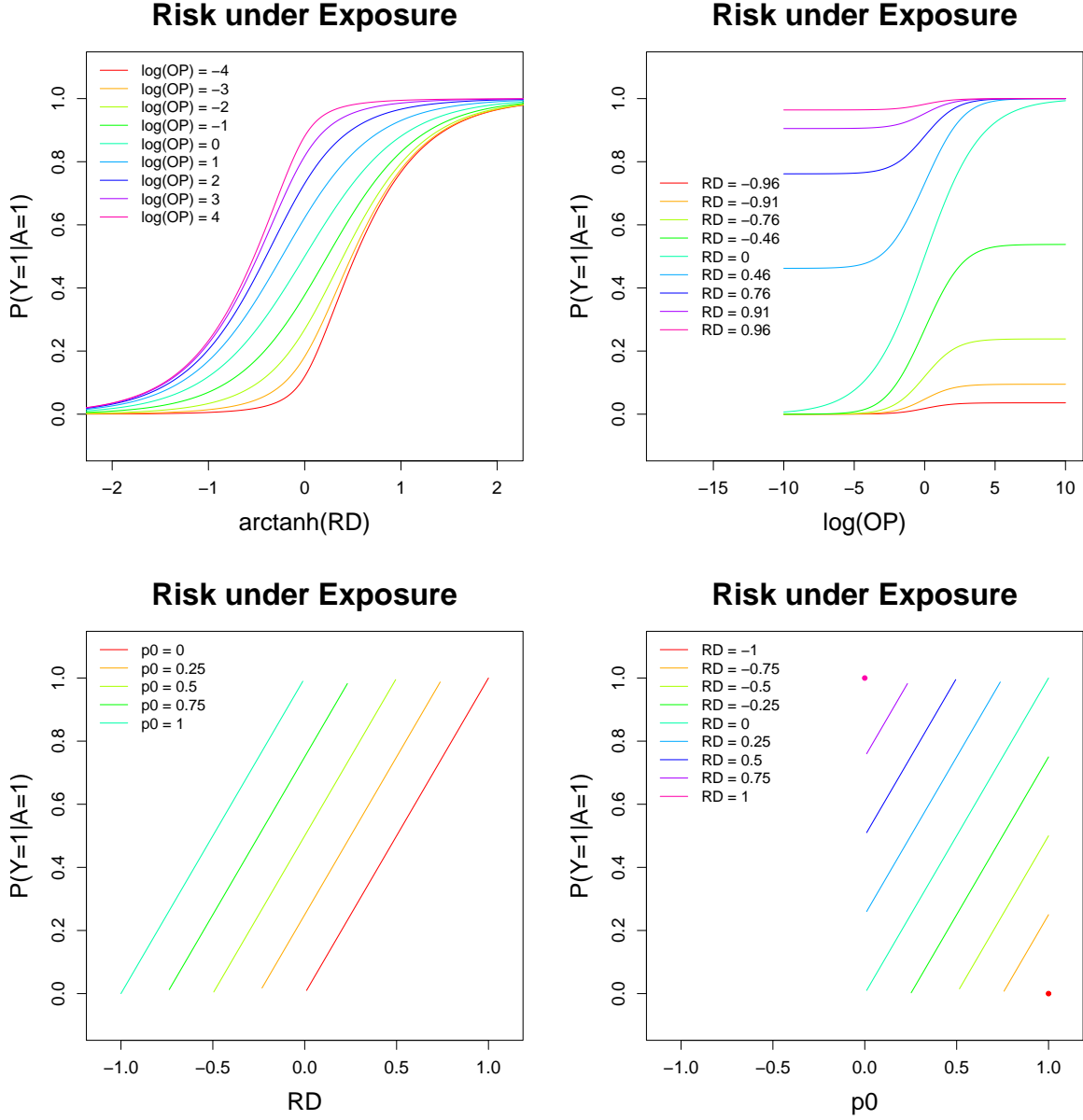


Figure 4: Fitted probabilities from the proposed risk difference model (upper panels) and the linear mean model (lower panels). For the proposed risk difference model, curves for fitted risks under non-exposure have the same shapes to those for fitted risk under exposure.

where

$$\omega_{eff}(V) = \frac{W}{e(V)(1-p_0)} \frac{E\left[\frac{Ap_A}{1-p_A} \mid V\right]}{E\left[\frac{p_A}{1-p_A} \mid V\right]}.$$

The model (2.1) is now given by

$$p_A = p_0(V) + \tanh(\theta(V))A.$$

The score function S_α is hence

$$S_\alpha = \frac{A\rho_\alpha(Y - p_A)}{p_A(1 - p_A)},$$

where $\rho \equiv \text{RD}(V) = \tanh(\theta(V))$ and $\rho_\alpha = \frac{\partial}{\partial \alpha} \tanh(\theta(V)) = (1 - \rho^2)W$.

To calculate the tangent space T_η , we note that it is the direct sum of three orthogonal spaces: $T_\eta = T_{p_0} \oplus T_{A|V} \oplus T_V$. Moreover, as α depends on the observed data distribution only through $p_{Y|A,V}$, the projection of the score function onto $T_{A|V} \oplus T_V$ equals zero. To calculate the tangent space T_{p_0} , we consider submodels centered on p_0 and indexed by a scalar δ (taking values in a neighborhood of zero) that are implied by $p_A = p_0(V) + A\rho + \delta h(V)$ for an arbitrary function $h(V)$. It is easy to see that this indeed implies a submodel obeying (2.1). The score of this submodel at $\delta = 0$ equals $h(V)(Y - p_A)/(p_A(1 - p_A))$; hence

$$T_{p_0} = \left\{ \frac{Y - p_A}{p_A(1 - p_A)} h(V) : h(V) \right\}.$$

To obtain the projection of S_α onto T_{p_0} , we need to find $h^*(V)$ such that,

$$E \left[\left\{ u(A, V)(Y - p_A) - \frac{h^*(V)(Y - p_A)}{p_A(1 - p_A)} \right\} \frac{h(V)(Y - p_A)}{p_A(1 - p_A)} \right] = 0, \text{ for all } h(V),$$

where $u(A, V) = A\rho_\alpha/(p_A(1 - p_A))$. Integrating out Y gives

$$E \left[h(V) \frac{(u(A, V)p_A(1 - p_A) - h^*(V))}{p_A(1 - p_A)} \right] = 0, \text{ for all } h(V).$$

Hence

$$h^*(V) = \frac{E \left[\frac{A\rho_\alpha}{p_A(1 - p_A)} \mid V \right]}{E \left[\frac{1}{p_A(1 - p_A)} \mid V \right]} = \frac{E \left[\frac{A(1 - \rho^2)}{p_A(1 - p_A)} \mid V \right]}{E \left[\frac{1}{p_A(1 - p_A)} \mid V \right]} W.$$

Consequently,

$$S^*(\dot{\alpha}, \dot{\eta}) = \frac{Y - p_A}{p_A(1 - p_A)} \left(A\rho_\alpha - \frac{E \left[\frac{A\rho_\alpha}{p_A(1 - p_A)} \mid V \right]}{E \left[\frac{1}{p_A(1 - p_A)} \mid V \right]} \right).$$

By analytic calculation, it is easy to verify that

$$S^*(\dot{\alpha}, \dot{\eta}) = \omega_{eff}(V)(A - e(V))(H(\dot{\alpha}) - p_0(V)),$$

where $\omega_{eff} = h^*(V)/(e(V)p_0(1 - p_0))$. ■

C Proof of Theorem 3.2

Proof. By Slutsky's theorem, it is easy to see that the probability limit of (3.5) is

$$U(\alpha; \alpha^*, \beta^*, \gamma^*; \omega) \equiv \omega(V) (A - e(V; \gamma^*)) (H(\alpha) - p_0(V; \alpha^*, \beta^*)) = 0. \quad (6.1)$$

Let $\dot{\beta}$ and $\dot{\gamma}$ be the true values of β and γ , respectively. When model (2.2) is correct, $(\alpha^*, \beta^*) = (\dot{\alpha}, \dot{\beta})$; when model (3.3) is correct, $\gamma^* = \dot{\gamma}$.

To show that $\hat{\alpha}_{DR}(\omega)$ is a RAL estimator of α in the union model characterized by (2.1) and either (2.2) or (3.3), we need to show that under this union model,

- (i) (6.1) is unbiased for estimating α ;
- (ii) there exists only one solution to the estimating equation

$$E[U(\alpha; \alpha^*, \beta^*, \gamma^*; \omega)] = 0. \quad (6.2)$$

The asymptotic linearity of $\hat{\alpha}_{DR}$ and the corresponding influence function formula then directly follow from Van der Vaart (2000, Theorem 5.21).

Proof of unbiasedness

Observe that under (2.1), if $\theta(V) = \log \text{RR}(V)$, then

$$\begin{aligned} E[H(\dot{\alpha})|V] &= E\left[Y e^{-A\dot{\alpha}^T W} \middle| V\right] \\ &= p_0(V)(1 - e(V)) + p_1(V) \left(\frac{p_0(V)}{p_1(V)}\right) e(V) \\ &= p_0(V); \end{aligned}$$

likewise if $\theta(V) = \text{arctanh}(\text{RD}(V))$, then

$$\begin{aligned}
E[H(\dot{\alpha})|V] &= E[Y - A \tanh(\dot{\alpha}^T W)|V] \\
&= p_0(V)(1 - e(V)) + (p_1(V) - (p_1(V) - p_0(V)))e(V) \\
&= p_0(V).
\end{aligned}$$

Similarly, one can show that $E[AH(\dot{\alpha})|V] = p_0(V)e(V)$.

Now suppose that (2.2) is correctly specified, then $(\alpha^*, \beta^*) = (\dot{\alpha}, \dot{\beta})$ and $p_0(V) = p_0(V; \dot{\alpha}, \dot{\beta})$. Hence

$$\begin{aligned}
E\{U(\dot{\alpha}; \alpha^*, \beta^*, \gamma^*; \omega)|V\} &= E\left[\omega(V)(A - e(V; \gamma^*))(H(\dot{\alpha}) - p_0(V; \dot{\alpha}, \dot{\beta})|V)\right] \\
&= \omega(V)\{E[AH(\dot{\alpha}) - Ap_0(V)|V] - e(V; \gamma^*)E[H(\dot{\alpha}) - p_0(V)|V]\} \\
&= 0.
\end{aligned}$$

On the other hand, if (3.3) is correctly specified, then $\gamma^* = \dot{\gamma}$ and $e(V) = e(V; \dot{\gamma})$. Hence

$$\begin{aligned}
&E\{U(\dot{\alpha}; \alpha^*, \beta^*, \gamma^*; \omega)|V\} \\
&= E[\omega(V)(A - e(V; \dot{\gamma}))(H(\dot{\alpha}) - p_0(V; \alpha^*, \beta^*)|V)] \\
&= \omega(V)\{E[AH(\dot{\alpha}) - e(V)H(\dot{\alpha})|V] - p_0(V; \alpha^*, \beta^*)E[A - e(V)|V]\} = 0.
\end{aligned}$$

Proof of existence and uniqueness of solution to (6.2)

Let

$$U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) \equiv \omega(V)(A - e(V; \gamma^*))(H(\theta) - p_0(V; \alpha^*, \beta^*))$$

where $H(\theta) = Y\{\exp(-A\theta)\}$ when $\theta(V) = \log(\text{RR}(V))$, and $H(\theta) = Y - A \tanh(\theta)$ when $\theta(V) = \text{arctanh}(\text{RD}(V))$. (In other words, $U(\theta; \alpha^*, \beta^*, \gamma^*; \omega)$ is (3.6) with $\alpha^T W$ replaced by θ .)

We now show that if either (2.2) or (3.3) is correctly specified, with probability 1, there is at most one value $\theta(v)$ solving

$$E \{U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v\} = 0. \quad (6.3)$$

Therefore, assuming a non-degenerate distribution for $p_0(V)$, there exists only one α such that $\theta(V) = \alpha^T W$ with probability 1.

For simplicity, for any given v , we write $e^*(v) = e(v; \gamma^*)$ and $p_0^*(v) = p_0(v; \alpha^*, \beta^*)$. Without loss of generality, we assume that with probability 1, $\omega(v) > 0$. If $\theta(v) = \log \text{RR}(v)$, then

$$\begin{aligned} & \frac{\partial}{\partial \theta} E [U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] \\ &= \frac{\partial}{\partial \theta} E [\omega(V)(A - e^*(V))(Y e^{-A\theta} - p_0^*(V)) | V = v] \\ &= -\omega(v) E [(A - e^*(v))AY e^{-A\theta} | V = v] \\ &= -C_1(v) e^{-\theta} < 0, \end{aligned}$$

where $C_1(v) = \omega(v)e(v)(1 - e^*(v))p_1(v)$ is a positive constant given v . We interchange the order of derivative and integration in the second step as for any value of θ , one can find a bounded open neighborhood (a, b) around it such that $(A - e^*(v))AY e^{-A\theta}$ is integrable. This establishes that there is at most one solution to (6.3).

To establish existence, we note that

$$\begin{aligned} & \lim_{\theta \rightarrow \infty} E [U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] \\ &= \lim_{\theta \rightarrow \infty} E [\omega(V)(A - e^*(V))(Y e^{-A\theta} - p_0^*(V)) | V = v] \\ &= \omega(v)(1 - e(v))(-e^*(v))(p_0(v) - p_0^*(v)) + \\ & \quad \omega(v)e(v)(1 - e^*(v))(-p_0^*(v)) \end{aligned} \quad (6.4)$$

$$\begin{aligned} &= \omega(v)p_0^*(v)(e^*(v) - e(v)) + \\ & \quad \omega(v)(1 - e(v))(-e^*(v))p_0(v). \end{aligned} \quad (6.5)$$

If (2.2) is correctly specified, then the first term in (6.4) vanishes, and the remaining term in (6.4) is negative; likewise if (3.3) is correctly specified, then the first term in (6.5) vanishes,

and the remaining term in (6.5) is negative. Hence if either (2.2) or (3.3) is correctly specified, we have

$$\lim_{\theta \rightarrow \infty} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] < 0.$$

As $\frac{\partial}{\partial \theta} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] < -C_1(v)$ for $\theta < 0$, we have

$$\lim_{\theta \rightarrow -\infty} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] = \infty.$$

By continuity, this establishes the existence of a solution to (6.3).

If $\theta(v) = \text{arctanh}(\text{RD}(v))$, then

$$\begin{aligned} & \frac{\partial}{\partial \theta} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] \\ &= \frac{\partial}{\partial \theta} E[\omega(V)(A - e^*(V))(Y - A \tanh(\theta) - p_0^*(V)) | V = v] \\ &= -\omega(v) E[(A - e^*(v))A(1 - \tanh^2(\theta)) | V = v] \\ &= -C_2(v)(1 - \tanh^2(\theta)) \\ &< 0, \end{aligned}$$

where $C_2(v) = \omega(v)e(v)(1 - e^*(v))$ is a positive constant given v . This establishes there is at most one solution to (6.3).

To establish existence, we note

$$\begin{aligned} & \lim_{\theta \rightarrow \infty} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] \\ &= \lim_{\theta \rightarrow \infty} E[\omega(V)(A - e^*(V))(Y - A \tanh(\theta) - p_0^*(V)) | V = v] \\ &= E[\omega(V)(A - e^*(V))(Y - A - p_0^*(V)) | V = v] \\ &= \omega(v)(1 - e(v))(-e^*(v))(p_0(v) - p_0^*(v)) + \\ & \quad \omega(v)e(v)(1 - e^*(v))(p_1(v) - 1 - p_0^*(v)) \end{aligned} \tag{6.6}$$

$$\begin{aligned} &= \omega(v)p_0^*(v)(e^*(v) - e(v)) - \\ & \quad \omega(v)(1 - e(v))e^*(v)p_0(v) + \\ & \quad \omega(v)e(v)(1 - e^*(v))(p_1(v) - 1). \end{aligned} \tag{6.7}$$

If (2.2) is correctly specified, then the first term in (6.6) vanishes, and the remaining term in (6.6) is negative; likewise if (3.3) is correctly specified, then the first term in (6.7) vanishes, and the remaining term in (6.7) is negative. Hence if either (2.2) or (3.3) is correctly specified, we have

$$\lim_{\theta \rightarrow \infty} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] < 0.$$

Likewise

$$\begin{aligned} & \lim_{\theta \rightarrow -\infty} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] \\ &= \lim_{\theta \rightarrow -\infty} E[\omega(V)(A - e^*(V))(Y - A \tanh(\theta) - p_0^*(V)) | V = v] \\ &= E[\omega(V)(A - e^*(V))(Y + A - p_0^*(V)) | V = v] \\ &= \omega(v)(1 - e(v))(-e^*(v))(p_0(v) - p_0^*(v)) + \\ & \quad \omega(v)e(v)(1 - e^*(v))(p_1(v) + 1 - p_0^*(v)) \end{aligned} \tag{6.8}$$

$$\begin{aligned} &= \omega(v)p_0^*(v)(e^*(v) - e(v)) - \\ & \quad \omega(v)(1 - e(v))e^*(v)p_0(v) + \\ & \quad \omega(v)e(v)(1 - e^*(v))(p_1(v) + 1). \end{aligned} \tag{6.9}$$

If (2.2) is correctly specified, then the first term in (6.8) vanishes, and the remaining term in (6.8) is positive; likewise if (3.3) is correctly specified, then the first term in (6.9) vanishes, and the remaining term in (6.9) equals $\omega(v)e(v)(1 - e(v))(1 + p_1(v) - p_0(v))$ and is hence positive. Hence if either (2.2) or (3.3) is correctly specified, we have

$$\lim_{\theta \rightarrow \infty} E[U(\theta; \alpha^*, \beta^*, \gamma^*; \omega) | V = v] > 0.$$

By continuity, this establishes existence of a solution to (6.3). ■

D Proof of Theorem 3.6

Proof.

If both the nuisance model (2.2) and the propensity score model (3.3) are correct, then

$(\alpha^*, \beta^*) = (\dot{\alpha}, \dot{\beta})$ and $\gamma^* = \dot{\gamma}$. Hence we have

$$\begin{aligned} & E \left[\frac{\partial U(\dot{\alpha}; \alpha, \beta, \gamma^*)}{\partial(\alpha^T, \beta^T)} \Big|_{(\alpha, \beta) = (\alpha^*, \beta^*)} \right] \\ &= E \left[\omega_{\text{eff}}(V)(A - e(V)) \left(-\frac{\partial p_0(V; \alpha, \beta)}{\partial(\alpha, \beta)} \right) + \right. \\ &\quad \left. \frac{\partial \omega_{\text{eff}}(V; \alpha, \beta, \dot{\gamma})}{\partial(\alpha, \beta)} (A - e(V)) (H(\dot{\alpha}) - p_0(V)) \Big|_{(\alpha, \beta) = (\alpha^*, \beta^*)} \right] = 0 \end{aligned}$$

and

$$\begin{aligned} & E \left[\frac{U(\dot{\alpha}; \alpha^*, \beta^*, \gamma)}{\partial \gamma} \Big|_{\gamma = \gamma^*} \right] \\ &= E \left[\omega_{\text{eff}}(V) \left(-\frac{\partial e(V; \gamma)}{\gamma} \right) (H(\dot{\alpha}) - p_0(V)) + \right. \\ &\quad \left. \frac{\partial \omega_{\text{eff}}(V; \dot{\alpha}, \dot{\beta}, \gamma)}{\partial \gamma} (A - e(V)) (H(\dot{\alpha}) - p_0(V)) \Big|_{\gamma = \gamma^*} \right] = 0. \end{aligned}$$

Following Theorem 3.2, the influence function of $\hat{\alpha}_{DR, \text{eff}}$ is $\tau^{-1}U(\dot{\alpha}; \dot{\alpha}, \dot{\beta}, \dot{\gamma})$, where $\tau = -E \left[\partial U(\alpha; \dot{\alpha}, \dot{\beta}, \dot{\gamma}) / \partial \alpha|_{\alpha = \dot{\alpha}} \right]$. From the discussion following Theorem 3.1, it is evident that the variance of $\hat{\alpha}_{DR, \text{eff}}$ attains the bound. ■

References

- Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity*, 43(1):59–74.
- L'Abbé, K. A., Detsky, A. S., and O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, 107:224–233.
- Lumley, T., Kronmal, R., and Ma, S. (2006). Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series*. Working paper 293. <http://biostats.bepress.com/uwbiostat/paper293/>. Last accessed on Jul 13, 2015.

- McNutt, L.-A., Wu, C., Xue, X., and Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American journal of epidemiology*, 157(10):940–943.
- Neutra, R. R., Greenland, S., and Friedman, E. A. (1980). Effect of fetal monitoring on cesarean section rates. *Obstetrics & Gynecology*, 55(2):175–180.
- Robins, J. M. and Rotnitzky, A. (2001). Comment on the Bickel and Kwon article, “Inference for semiparametric models: Some questions and an answer”. *Statistica Sinica*, 11(4):920–936.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.
- Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins, 3rd edition.
- Rotnitzky, A. and Vansteelandt, S. (2014). Double-robust methods. In Fitzmaurice, G., Kenward, M., Molenberghs, G., Tsiatis, A., and Verbeke, G., editors, *Handbook of Missing Data Methodology*. Chapman & Hall/CRC Press.
- Tchetgen Tchetgen, E. J. (2013). Estimation of risk ratios in cohort studies with a common outcome: a simple and efficient two-stage approach. *The International Journal of Biostatistics*, 9(2):251–264.
- Tchetgen Tchetgen, E. J., Robins, J. M., and Rotnitzky, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*, 97(1):171–180.
- Van Der Laan, M. J., Hubbard, A., and Jewell, N. P. (2007). Estimation of treatment effects in randomized trials with non-compliance and a dichotomous outcome. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):463–482.
- Van der Laan, M. J. and Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer Science & Business Media.

- Van der Laan, M. J. and Rose, S. (2011). *Targeted learning: causal inference for observational and experimental data*. New York, NY: Springer Publishing Company.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.
- Williamson, T., Eliasziw, M., and Fick, G. H. (2013). Log-binomial models: exploring failed convergence. *Emerging themes in epidemiology*, 10(1):14.